**Dietmar G. Schrausser**
Institut für Psychologie der Universität Graz

# Exakte Verfahren oder Asymptotische Approximation?

## Einleitung

Sollen statistisch abgesicherte Aussagen über Lageparameter von Merkmalen in einer oder mehreren Stichproben getroffen werden, so geht man üblicherweise nach dem Konzept der Asymptotischen Approximation (AP) vor. Die dazugehörige Annahme über Verteilungsformen von Kennwerten $\theta$ aus Stichproben einer Population stellt allerdings hohe Voraussetzungen an das Datenmaterial: Die Stichprobendaten müssen (a) einer zumindest symmetrischen, wenn nicht normalen Verteilung folgen, müssen (b) homogen in ihren Varianzen, (c) genügend groß und (d) durch einen Zufallsprozeß zustande gekommen sein. Sind diese Bedingungen nicht erfüllt, so sind Wahrscheinlichkeitsaussagen über das Vorkommen von Parametern bzw. Parameterdifferenzen logisch nicht mehr definiert. Dieses Faktum ist von entscheidender Bedeutung in der psychologischen Forschung, da man es hier nur selten mit Daten zu tun hat, die den Ansprüchen der asymptotischen Vorgangsweise voll gerecht werden. Alternative Vorgangsweisen sind einerseits nonparametrische Verfahren bzw. Korrekturalgorithmen, durch deren Anwendung jedoch die Aussagekraft der Ergebnisse relativiert wird, andererseits sog. Exakte Verfahren bzw. Permutationstests.

Solche Permutations- oder Randomisationstests wurden erstmals von Sir Ronald A. Fisher im Jahre 1935, basierend auf eigenen Ausführungen zu Experimenten im Agrarbereich (Fisher, 1926) und Arbeiten von Neyman (1923) erwähnt. Fisher's Konzept des Randomisierungstests wurde von Pitman (1937a,b, 1938) und Welch (1937) weiterentwickelt und auf komplexere Probleme (wie z.B. das 2-Stichprobenproblem vgl. Schrausser, 1996) übergeführt. Diese Tests liefern auch dann valide Signifikanztests, wenn man (a) aufgrund der unzureichenden Güte der Stichprobendaten nicht auf Populationsparameter rückschließen kann (reiner Permutationstest P) und bzw. oder (b) keine Zufallsstichprobe aus einer definierten Population zur Verfügung hat (Permutationstest im Randomisierungsmodell oder Randomisierungstest PR). Der erste Übersichtsaufsatz über die Verwendung von Permutationstests in der Psychologie stammt von Moses (1952). Verwandte Verfahrensweisen sind das Resampling- bzw. Bootsrap-Verfahren Bt (Efron, 1979) und die Jackknife Technik Jk (Quenouille, 1949). Die Unterschiede zwischen Bootstrap und Jackknife sind lediglich von geringem Ausmaß.

## Zur Methode: Bootstrap vs. Permutationstest

Der Unterschied der Permutationsmethode zu der Bootstrap Methode besteht darin, daß (a) bei der Permutationsmethode oder dem Permutationstest Kennwerte von 2 oder mehreren Stichproben verglichen werden (analog zu t-Test, U-Test o.ä.), während bei der Bootstrap-Methode primär der Kennwertfehler (z.B. von AM, Md, s) einer Stichprobe geschätzt wird. Weiters handelt es sich (b) bei der Bootstrap Methode um eine Erzeugung aller möglichen Stichproben aus einer vorliegenden Stichprobe, bei der Permutationsmethode hingegen um die Erzeugung aller möglichen Versuchspläne aus einem vorliegenden Versuchsplan.

Bei der Bootstrap Methode kommen die Elemente in jedem Resampel nur so oft vor, wie sie in der ursprünglichen Stichprobe vorkamen. In den Termini der Kombinatorik gesprochen werden bei der Bootstrap-Methode Variationen von $n$ Elementen zur Klasse $n$ bzw. $k$ mit Wiederholung gebildet, ${}^{W}V(n;k)={}^{W}V_n^k = n^k$. Bei der Permutationsmethode handelt es sich um Variationen von $n$ Elementen zur Klasse $k=n$ ohne Wiederholung, $V(n;k) = V_n^k = \binom{n}{k}k! = \frac{n!}{(n-k)!}$. Von den $n$ Elementen sind je $p_1$, $p_2$, ..., $p_i$ Elemente unter sich gleich (Permutation innerhalb der Gruppen). Die Anzahl der möglichen Permutationen $M$ zwischen den Gruppen, folgt aus $M = p_1, p_2, \ldots P(n) = \frac{n!}{p_1! \cdot p_2! \ldots p_i}$. Bis zu welchem Grade die beiden Methoden ineinander überführbar sind, wäre zu diskutieren.

Nullhypothesen lassen sich als Invarianzen der gemeinsamen Verteilungen der Elemente einer Stichprobe $f(x)$ gegenüber bestimmten Permutationsgruppen $Q_n$, Teilmengen aller möglichen Permutationen $S_n$, darstellen (vgl. Scheffe, 1959). Die Definition der Nullhypothese ist (nach Willmes, 1987) $H_0$: $f(x) = f(\pi x) \quad \forall\ \pi \in Q_n\ ,\ x \in R^n$ Wobei $f(\pi x)$ die Funktion der permutierten Elemente und $R^n$ die Elemente der Stichprobe bezeichnet. Das empirische Signifikanzniveau $p(x)$ ist, definiert nach Cox & Hinckley (1974) und

Kempthorne (1976) $p(x) = p(\theta_{(x)} \geq \Theta_{(x)} | H_0), 0 \leq p \leq 1$, **d**ie Wahrscheinlichkeit, bei $H_0$ Kennwerte $\theta_{(x)}$ zu beobachten, die größer oder gleich dem Kennwert $\Theta_{(x)}$ der experimentellen Anordnung sind. Grundlegende mathematische Ansätze zur Ableitung und Konstruktion von Invarianz-Permutationstests gibt Pyhel (1978, 1980a, b). Eine Zusammenfassung der mathematischen Herleitung ist bei Willmes (1987, S.37ff) zu finden.

## Beispiel: Der Permutationstest beim unabhängigen t-Test Paradigma

5 Personen werden per Zufall 2 Bedingungen oder Treatments zugeordnet: Bedingung A, $n_1$=3 Personen; Bedingung B, $n_2$=2 Personen. Aufgrund der wenigen Fälle ($n_1$+$n_2$=$n$=5) können keinerlei Annahmen über die Populationsverteilung getroffen werden, was eine sinnvolle Verwendung asymptotischer Verfahren ausschließt.

**Tabelle 1**
Experimentalanordnung der Personen, entsprechende Arithmetische Mittel und der Kennwert $\Theta$=$t$

| A | B |
|---|---|
| 18 | 6 |
| 30 | 12 |
| 54 | |
| 34 | 9 |
| 1.81 | |

Mit diesen Daten (Tab. 1) wird nun z.B. ein unabhängiger $t$-Test durchgeführt ($t$= 1.81). Es sei an dieser Stelle bemerkt, daß nicht notwendigerweise ein solcher $t$-Test verwendet werden muß, um einen Lageparameter-Unterschied zu charakterisieren. Edgington (1980, S. 44) erläutert, daß jeder andere Test ebenfalls zulässig ist, sofern die Ergebnisse eine überzufällige Übereinstimmung, sprich Korrelation aufweisen: " Two test statistics are equivalent if and only [!] if they are perfectly monotonically correlated over all data permutions in the set." Es wäre also in diesem Falle die einfache Mittelwertsdifferenz $\bar{x}_A - \bar{x}_B$ oder schlicht die Rohwertsumme $\Sigma x$ einer Treatmentgruppe als Kennwert $\theta$ denkbar. Sollen Zusammenhangshypothesen geprüft werden, so kann ein Korrelationskoeffizient $r$ berechnet werden. Die Wahl des jeweiligen Kennwertes $\theta$ kann je nach Fragestellung erfolgen und ist an keine weiteren Bedingungen gebunden.
Mit dieser Datenanordnung werden nun die $M = \frac{5!}{3!\cdot 2!} = \frac{120}{12} = 10$ möglichen Permutationen der Daten zwischen den Bedingungen **A** und **B** durchgeführt. Dies entspricht allen möglichen unterschiedlichen Anordnungen der Personen in dem Versuchsplan (s.o.). Für jede dieser Anordnungen wird wiederum der gewählte Kennwert $\theta$=$t$ berechnet (s. Tab. 2).

**Tabelle 2** Mögliche Permutationen der Daten zwischen den Bedingungen, entsprechende Arithmetische Mittel und die Kennwerte $\theta$=$t$ bzw. $\Theta$=$t$

| A | B | A | B | A | B | A | B | A | B |
|---|---|---|---|---|---|---|---|---|---|
| 18 | 30 | 6 | 18 | 18 | 6 | 8 | 54 | 6 | 18 |
| 6 | 54 | 30 | 54 | 30 | 54 | 30 | 12 | 12 | 30 |
| 12 | | 12 | | 12 | | 6 | | 54 | |
| 12 | 42 | 16 | 36 | 20 | 30 | 18 | 33 | 24 | 24 |
| -3.00 | | -1.22 | | -0.52 | | -0.33 | | 0.00 | |

| A | B | A | B | A | B | A | B | A | B |
|---|---|---|---|---|---|---|---|---|---|
| 18 | 30 | 18 | 6 | 6 | 18 | 12 | 6 | 18 | 6 |
| 6 | 12 | 12 | 30 | 30 | 12 | 30 | 18 | 30 | 12 |
| 54 | | 54 | | 54 | | 54 | | 54 | |
| 26 | 21 | 28 | 18 | 30 | 15 | 32 | 12 | 34 | 9 |
| 0.25 | | 0.52 | | 0.83 | | 1.22 | | 1.81 | |

Experimentalanordnung

Das Signifikanzniveau $p$ ergibt sich aus dem Verhältnis des Kennwertes $\Theta$ aus der Experimentalanordnung zu den Kennwerten $\theta$, die für jede permutierte Anordnung errechnet wurden. Zu beachten ist, daß der Kennwert $\Theta$ aus der Experimentalanordnung in die Menge der Kennwerte der permutierten Anordnungen mit eingeht. Abhängig davon, ob die Fragestellung 1- oder 2-seitig erfolgte, werden entweder der Kennwert mit Vorzeichen oder der absolute Wert verwendet. In unserem Falle nehmen wir eine 2-seitige Fragestellung an: 2 der 10 Kennwerte oder absoluten t-Werte ($t$=1.81, $t$=-3.00) sind größer oder gleich $t$=1.81, jenem $\Theta$, das aus der experimentellen Anordnung resultierte. Das Verhältnis von 2 (Werte größer oder gleich Kennwert $\Theta$) zu der Anzahl von Permutationen $M$=10 oder errechneten Kennwerten $\theta$ ergibt das 2-seitige Signifikanzniveau $p = \frac{\theta \geq \Theta}{M} = \frac{2}{10} = 0.20$, $|t|$ ist am 2-seitigen 20% Niveau signifikant.

Diese grundlegenden Schritte zur Durchführung eines Permutationstests bleiben für alle Versuchspläne (Meßwiederholungsdesigns, faktorielle Pläne im uni- oder multivariaten Fall, Einzellfalluntersuchungen, Zeitreihen usw.) prinzipiell gleich. (vgl. Edgington, 1995, Schrausser & Scambor, 1995). Eine Darstellung der Vorgangsweise bei abhängigen Stichproben gibt Scambor (1997).

## Diskussion und Ausblick

Der Kardinalvorzug von exakten Verfahrensweisen (P bzw. Bt) ist das Wegfallen von Verteilungsvoraussetzungen an die Stichprobendaten und somit die Möglichkeit, auch aus sehr geringen Stichprobenumfängen noch logisch begründbare Aussagen über deren statistische Bedeutsamkeit hinsichtlich eines Kennwertes zu treffen. Darüber hinaus kann man einen beliebigen Kennwert zur Charakterisierung eines Merkmales relativ frei wählen. Die Einsatzbereiche sind vorwiegend dort zu sehen, wo man keinen großen Einfluß auf die Datenverteilungen und experimentellen Bedingungen nehmen kann. Dies ist in klinischen Untersuchungen genauso der Fall wie in quasiexperimentellen Designs, Metaanalysen oder vorexperimentellen Studien.

Kann das Wissen über die Funktionsweise und die Einsatzgebiete von Exakten Verfahren einem breiteren Personenkreis zugänglich gemacht werden, so könnte sich diese Vorgangsweise der statistische Inferenz durchaus als die Methode erster Wahl etablieren.

## Literatur


Cox, D. R., Hinckley, D.V. (1974). Theoretical statistics. London: Chapman and Hall.

Edgington, E. S. (1980). Randomization tests. New York: Marcel Dekker.

Edgington, E. S. (1995). Randomization tests (3nd ed). New York: Marcel Dekker.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. The annals of statistics, 1, 1-26.

Fisher R. A. (1926). The arrangement of field experiments. Journal of Ministry of Agriculture, 33, 503-513.

Franck, W. E. (1986). P-values for discrete test statistics. Biometrical Journal, 28, 403-406.

Kempthorne, O. (1976). Of what use are tests of significance and tests of hypothesis. Communication in Statistics - Theory and Methods. A5, 763 - 777.

Moses, L. E. (1952). Nonparametric statistics for psychological research. Psychological Bulletin, 49, 122-143.

Neyman, J. (1923). Sur les applications de la theorie des probabilites aux experience agricoles: Essay de principes. Roczniki Nank Polniczek, 10, 1-51.

Pitman, E. J. G. (1937a). Significance tests which may be applied to samples from any populations. Journal of the Royal Statistical Society Series B, 4, 119-130.

Pitman, E. J. G. (1937b). Significance tests which may be applied to samples from any populations II: The correlation coefficient test. Journal of the Royal Statistical Society Series B, 4, 225-232.

Pitman, E. J. G. (1938). Significance tests which may be applied to samples from any populations. III. The analysis of variance test. Biometrika, 29, 322-335.

Pyhel, N. (1978). Beiträge zur Theorie der Permutationstests mit Anwendung in einem linearen Modell. Dissertation. Mathematisch-Naturwissenschaftliche Fakultät, RWTH Aachen, Aachen.

Pyhel, N. (1980a). Distribution-free r sample tests for the hypothesis of parallelism of response profiles. Biometrical Journal, 22, 703-714.

Pyhel, N. (1980b). Zur Anwendung von Permutationstests in mehrfaktoriellen Versuchsplänen. In W. Köpcke, & K. Überla (Hrsg.), Biometrie - heute und morgen. Interregionales Biometrisches Kolloquium, 350-358. Berlin: Springer.

Quenoille, M. H. (1949). Approximate tests of correlation in time series. Journal of the Royal Statistical Society, Series B, 11, 68-84.

Scambor, C. (1997). Permutationsverfahren in Einzelfallstudien: Grundlagen, Anwendungen und Teststärke. Diplomarbeit. Institut für Psychologie, Karl-Franzens Universität, Graz.

Scheffe, H. (1959). The analysis of variance. New York: Wiley.

Schrausser, D. G. (1996). Permutationstests: Theoretische und praktische Arbeitsweise von Permutationsverfahren beim unverbundenen 2-Stichprobenproblem. Diplomarbeit. Institut für Psychologie, Karl-Franzens Universität, Graz.

Schrausser, D. G., & Scambor, C. (1995). Randomisierungstests. Eine Einführung in die Methode, unter besonderer Berücksichtigung ihrer Bedeutung in computerunterstützter Datenanalyse und Einzelfall- Versuchsplänen. Institut für Psychologie, Karl-Franzens Universität, Graz.

Welch, B. L. (1937). On the z-test in randomized blocks and Latin squares. Biometrika, 29, 21-52.

Willmes, K. (1987). Beiträge zu Theorie und Anwendung von Permutationstests in der Uni- und Multivariaten Datenanalyse. Diss.; Fachbereich I - Psychologie, Universität Trier, Trier.